Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Use Cases for Visualizing Uncertain Computer Networks	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,4555 Overlook Avenue SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
IEEE VisWeek Workshop on Working with Uncertainty, 23-28 Oct 2011, Providence, RI.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**2**	

Use Cases for Visualizing Uncertain Computer Networks

Jonathan W. Decker* Mark A. Livingston† Stephen Russell‡ Paul Hyden§

Naval Research Laboratory

Index Terms—Network visualization, uncertainty visualization, graph layout

1 Problem Overview and Background

Network administration has become an extremely complex task. As wireless devices have become more commonplace, along with varied resources such as web servers, storage devices, autonomous sensors, and printers, the network becomes more dynamic and it becomes harder to keep track of all the machines and users present. This becomes even more pressing when a user engages in deceptive practices for the purposes of launching an attack on the network resources. Such users deliberately obfuscate their identity, creating a level of uncertainty about the network. Most networking tools characterize trends in variables of interest with assorted two-dimensional statistical graphics and node-link diagrams. Users are often forced to navigate through various contextual combinations during the process of inspection, investigation, or hypothesis testing. While network monitoring and exploration tools abound, few visualize uncertainty in a complex network administration task.

There are clear benefits to having an accurate map of a given computer network. Applications of a network map include capacity analysis, maintenance and operation, and empirical threat detection. One way to determine the layout of a network is to actively probe each network component using information gathering packets. Unfortunately, there is no guarantee that a component will act accordingly when it receives a specific packet. A component could choose to dismiss the request, or worse, it could return bad information. Moreover, it can be expected that no detailed information can be obtained beyond a proxy server, and so the structure of a remote subnet will be completely unknown to the source machine. Passive monitoring suffers from many of the same problems as actively probing, but may minimize defensive mechanisms intended to prevent disclosure.

Regardless of the approach, network design patterns are often deliberately obfuscated for security reasons to the point of introducing deception and denial. Thus, in practice, a given network can be assumed to be stochastic and Byzantine. Therefore, an analytical approach is required to determine additional information, and this information would be associated with a degree of confidence. Each attribute gathered to assemble a network model is paired with some confidence value. If a second model gathers overlapping information, some attributes will begin to accumulate multiple confidence factors. However, the deception practices noted above could easily lead to inaccurate attribution of the certainty of data gathered or synthesized by any particular model. Further, there is implicit dependency in networks, which means that uncertainty propagates through the network analysis, further complicating visual tools to represent uncertain data.

A data-centric network view allows a user to focus on statistics for a selected network node, while a topology view shows cluster or network activity [7]. Dynamic networks need a temporal dimension, but this may be represented in glyphs rather than the dimension [5]; this work focused on "ego networks" (node of interest and its immediately adjacent nodes) and scalability, but not uncertainty. Comparing ego networks was done by a "bullseye" visualization using polar mapping of graph attributes in which graphical parameters separated node types. Parallel coordinates were used to visualize node uncertainty, with fisheye filters, coordinated multiple views, and brushing to help explore the space. Have Green [8] used an adjacency matrix to support uncertainty in queries on graphs. The Query Graph Visualizer [2] showed node-link diagrams with nodes representing queries and edges weighted by the degree of relationship between queries.

Zuk and Carpendale [9] analyzed a number of examples of uncertainty visualization in light of perceptual and cognitive principles. We add some observations that reflect these principles. Intensity in digital displays is perceived as continuous, but line width is too small in size to have the resolution to be perceived as continuous. Shape is inherently discrete. Our end users (network administrators) have strong mathematical backgrounds but not necessarily experience with reasoning under uncertainty. They monitor networks in which users behave as expected or (either unintentionally or intentionally) in ways that are disruptive. Response time to disruptive behavior is paramount.

2 Use Cases

We discuss two use cases in the area of network administration, not to limit the scope of the problem or the visualizations, but rather to present scenarios with different goals and types of uncertainty.

2.1 Network Structure and Membership

Understanding what devices are on a network is an easy task when the network is small, but the problem can quickly grow as the physical locations become more distributed and users are given authority to connect new devices. The latter case is common in the administration of a publicly-accessible wireless network. Our domain analysis named three entities on which uncertainty may exist in various forms: nodes (actors on the network), edges (connections between nodes), and subnets. Uncertainty exists for all these types. Nodes have a variety of categorical (e.g. type), discrete (number of users or peripherals), and continuous variables (bandwidth, trust, et al.). Similarly, edges have categorical properties such as connection (existence) and communication protocols used. A discrete measure would be the number of sessions in which it is participating. Continuous measures include latency and bandwidth. Figure 1 shows the mappings offered to our subject matter experts. Our mappings of the above variables reflect the principles stated above: node type with shape, number of users by border width, trust by fill intensity, number of sessions with chevrons, unresponsiveness with a box glyph and the time passed with saturation of the box, and latency with the density of arrow glyphs along the edge. One can see that the glyphs sometimes overlap. We do not explicitly avoid this, owing to the expected rarity of complete occlusion of a glyph that is vital to understanding. We do, however, ensure that the chevron glyph be in front, since it is likely to be more rare and is less dependent on its identical neighbors than the arrow glyph.

*email: jonathan.decker@nrl.navy.mil
†email: mark.livingston@nrl.navy.mil
‡email:stephen.russell@nrl.navy.mil
§email: paul.hyden@nrl.navy.mil

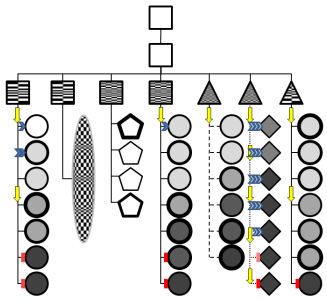

Fig. 1. We selected graphical attributes matched in continuous versus discrete nature to uncertainty variables.

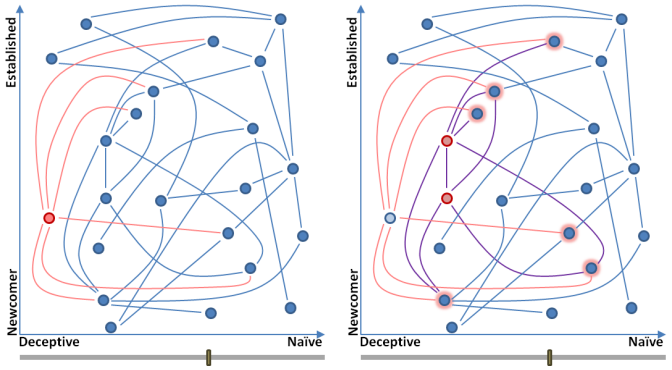

Fig. 2. *Left*: a node-link diagram is built in a scatterplot of intent and history (vertical). The red node is marked as suspected of deceptiveness. *Right*: Two nodes are fully connected (purple) to nodes in communication with the suspected node, raising suspicions about the behavior of these two (red) nodes. Note that neither communicates directly to the originally-suspected node.

2.2 Network Disruption

Analytical tools are beginning to emerge that enable an administrator to analyze network traffic and potential threats, as well as visualize the results or alerts [4, 6]. We describe a use case in which these analytical tools can be helpful in a visual representation of the network. Our subject matter experts divide the analytical metrics about actors on a network into two broad categories: intent and history. The first dimension denotes a measure that separates malicious behavior from other actions that can disrupt network performance through (for example) honest mistakes or naïve applications. The second dimension denotes any variable with a temporal dimension that can be used to mitigate the understanding of an alleged bad behavior. These two dimensions become the domain of a scatterplot. With the space defined, we now need to determine what content that domain has. A difference for our application from previous work is that the ego network is necessary, but often not sufficient to diagnose the source of the network disruption. Thus we use the confidence score of the intent to determine which nodes will be shown in the scatterplot. This reflects the uncertainty in the intent variable and gives the user control (scrollbar in Figure 2) of the number of nodes for which the problem diagnosis is considered.

By applying analytical tools such as Probabilistic Similarity Logic [1] (PSL), we can assist the administrator in reasoning about what is happening in this network. If we look for similarity of connections, we can understand what is happening. Figure 2 (left) shows the ego network of the suspected (red) node; this is a case in which the ego network is not sufficient. We apply PSL to see what other nodes are communicating with the ego network. Figure 2 (right) shows that two other nodes are fully connected to the same set of nodes; these nodes could be controlling the disruptive event and casting blame onto the initially-suspected member of the network.

3 Discussion

Open-source tools like Wireshark `http://www.wireshark.org/` or Nagios `http://www.nagios.org/` provide a GUI to help a user understand captured network traffic (a passive listening approach). A stream of packets captured from a network interface are listed chronologically, as they arrive. The entries in the list are highlighted by color, using protocol and other metrics (e.g. corrupted packets). The user can filter by protocol, TCP stream, and other keywords. The user can also click on any packet to see detailed information about it in a seperate panel. For a more visual representation of this information, there is a multi-view application called Cascade Pilot `http://www.riverbed.com/us/products/cascade/cascade_pilot.php`. This processes information in network traces to create a series of views. New views can be created interactively by selecting elements within the current views and drilling down to understand the traffic of specific hosts. The views include bar charts, line charts, pie charts, and a circular IP or MAC address conversions graph. Cascade Pilot works in conjunction with Wireshark to view the details of each packet. Similarly, products such as HP OpenView provided node-link diagrams with basic information about a known network's structure.

What we have presented to this point is a preliminary design; we are beginning the iterative process of refinement with our subject matter experts. One irregularity we note in the visual design is that we created intensity mappings with high brightness implying a high value and other mappings with high brightness implying a low value. Consistency in this aspect of the visual representations would likely serve the users better. A natural concern with the resulting layout is that edges could cross arbitrarily and create a cluttered graph. We plan to apply edge bundles [3] to mitigate this problem. Both the grouping of edges and the curving of edges around other nodes will help reduce the clutter in the layout of the graph created through the scatterplot.

Our central thesis is that a tool in the spirit of visual analytics will make it easier for network administrators to reason about the analytical outputs about a network when explicitly shown representations of the uncertainty associated with them. As noted in the network disruption use case, since the propagation of uncertainty can be a critical element in real scenarios, we feel that uncertainty visualization can assist in the analytical process.

References

[1] M. Bröcheler, L. Mihalkova, and L. Getoor. Probabilistic similarity logic. In *Proceedings of* 20^{th} *Conference on Uncertainty in Artificial Intelligence*, July 2010.

[2] D. H.-L. Goh, A. Y. K. Chua, C. S. Lee, and B. Luyt. Query graph visualizer: A visual collaborative querying system. In *First International Conference on the Applications of Digital Information and Web Technologies*, pages 78–83, Aug. 2008.

[3] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, September/October 2006.

[4] F. Mansmann, F. Fischer, D. A. Keim, S. Pietzko, and M. Waldvogel. Interactive analysis of netflows for misuse detection in large ip networks. In *DFN-Forum Kommunikationstechnologien*, pages 115–124, 2009.

[5] L. Shi, C. Wang, and Z. Wen. Dynamic network visualization in 1.5D. In *IEEE Pacific Visualisation Symposium*, pages 179–186, Mar. 2011.

[6] Y. Shi, Y. Tian, G. Kou, Y. Peng, and J. Li. *Network Intrusion Detection*, chapter 15, pages 237–241. Advanced Information and Knowledge Processing, Part 3. Springer, 2011.

[7] C. Waters, J. Howell, and T. J. Jankun-Kelly. CluVis: Dual-domain visual exploration of cluster/network metadata. In 45^{th} *Annual ACM Southeast Regional Conference*, pages 272–276, Mar. 2007.

[8] P. C. Wong, G. C. Jr., H. Foote, P. Mackey, and J. Thomas. Have Green – a visual analytics framework for large semantic graphs. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 67–74, Oct. 2006.

[9] T. Zuk and S. Carpendale. Theoretical analysis of uncertainty visualizations. In *Visualization and Data Analysis, Part of SPIE-IS&T Electronic Imaging, SPIE Vol. 6060*, Jan. 2006.